



DATA FUSION

CONTEST 2023

Meetup #2

11.03.2023

Welcome!

1. Промежуточные итоги

2. Companion

- Победители
- Выступления ++

3. Турнир

- Победители
- Гайды, patchnotes

4. “Референдум”

Track на **ODS.AI**:
Data Fusion Contest 2023



DATA

FUSION

CONTEST

Часть 1 — итоги

2023

Data Fusion Contest 2023



Задача
Атака

🏆 1 000 000 РУБ

👤 10

😊 10

🔥 3

+1

Задача 1 "Атака"

На что способен Adversarial ML в атаках на модели с данными транзакций? Попробуйте обойти, исказить и атаковать предложенную вам модель

👤 182 💎 505

February 7

April 2



Задача
Защита

🏆 1 000 000 РУБ

👤 5

💙 2

🏆 2

Задача 2 "Защита"

Можно ли защититься от современных атак на ML модель, не сильно потеряв в ее качестве? Обучите новую модель, устойчивую к Adversarial ML атакам

👤 110 💎 771

February 7

April 2



Специальная номинация
Companion

🏆 1 000 000 РУБ

За лучшие публичные решения

💙 8

🏆 3

👁 2

Специальная номинация "Companion"

Публикуй свои решения, собирай реакты от участников и получи денежный приз

👤 77 💎 11

February 7

April 2

Призы соревнования

	Атака	Защита
1.	125k	125k
2.	75k	75k
3.	50k	50k
с.	50k	50k

Награждение по итогам
первого турнира (сегодня)

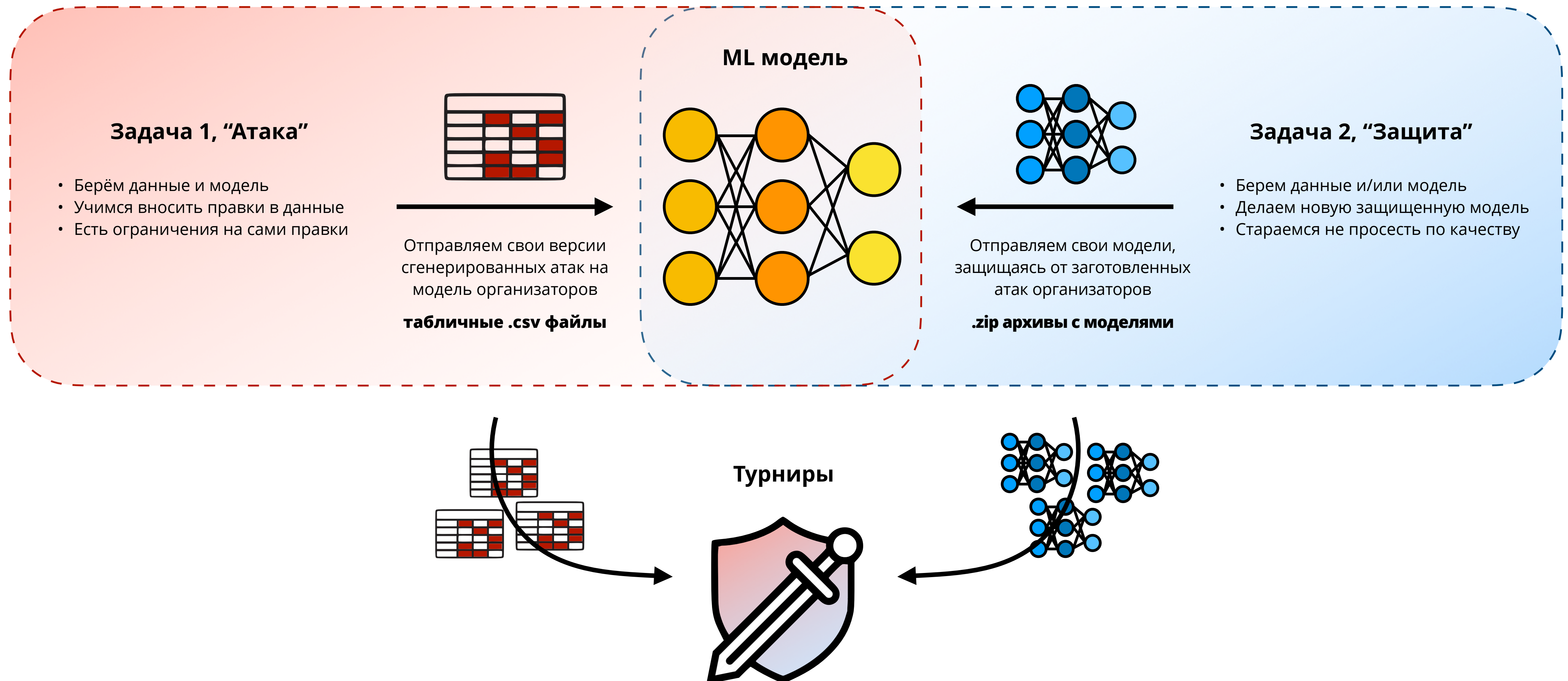
$\Sigma=600k$

Награждение по итогам
финального турнира (в апреле)

$\Sigma=1400k$

	Атака	Защита
1.	250k	250k
2.	150k	150k
3.	100k	100k
4.	75k	75k
5.	75k	75k
с.	50k	50k

Recap задач Data Fusion Contest 2023

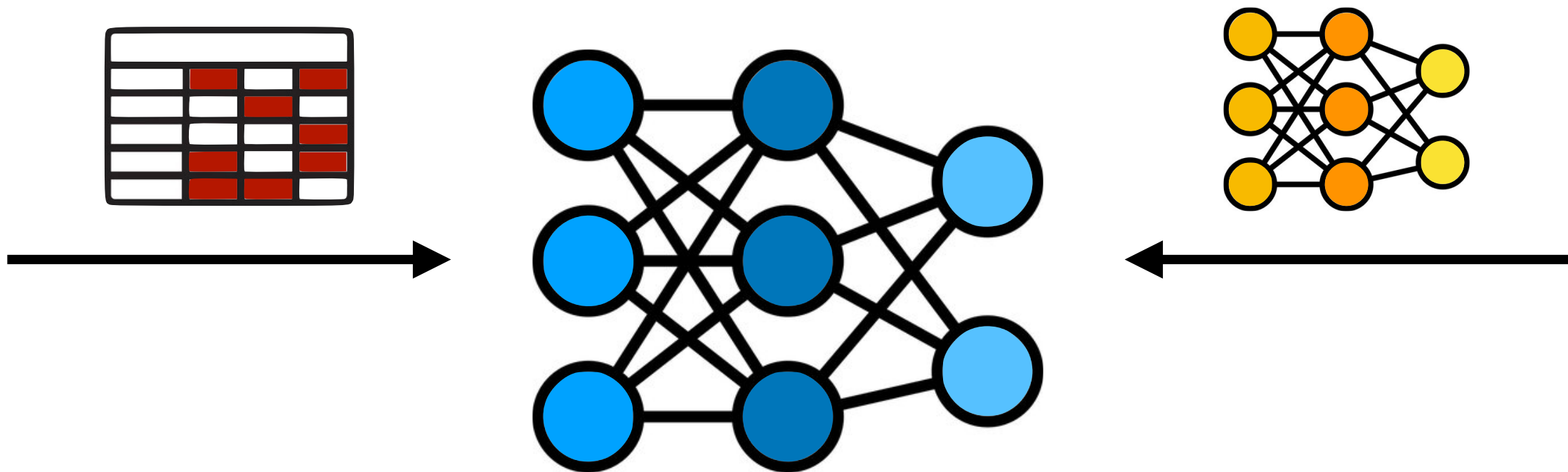


023.M03, ПИКТ В ЦВЕТЕ

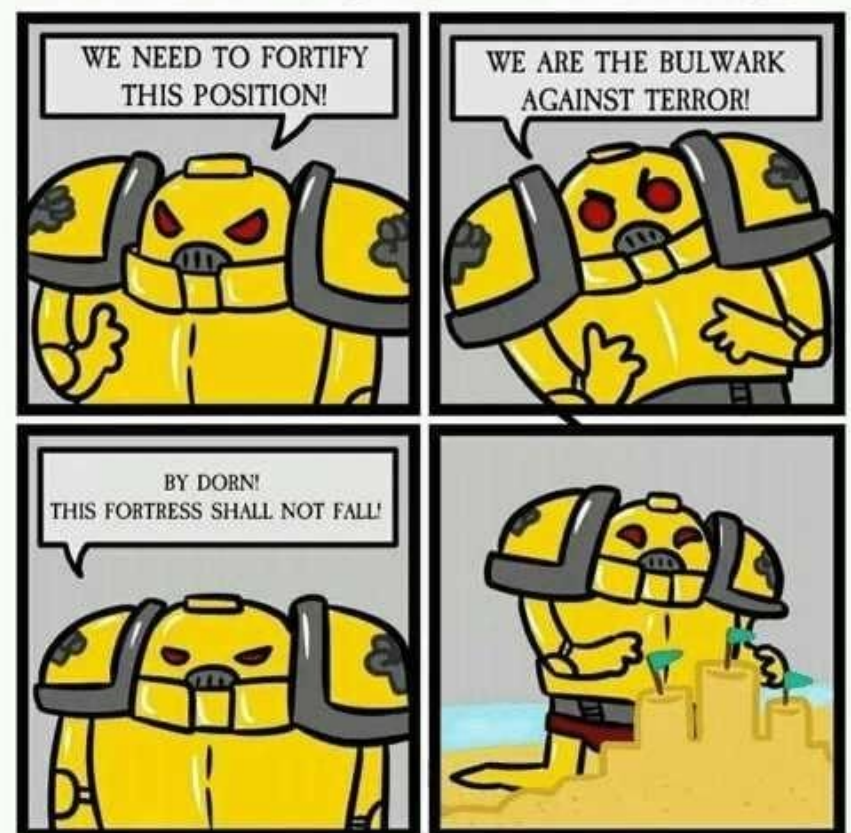
Задача 1, "Атака"



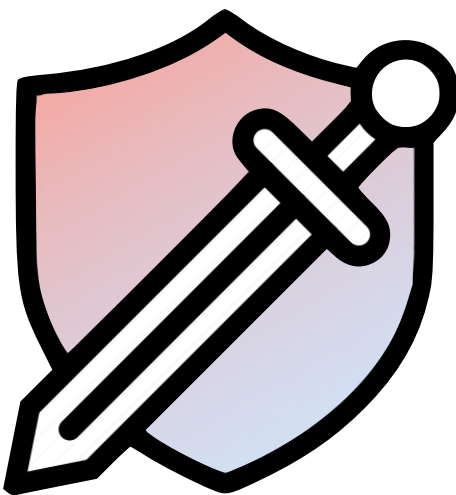
ML модель



Задача 2, "Защита"

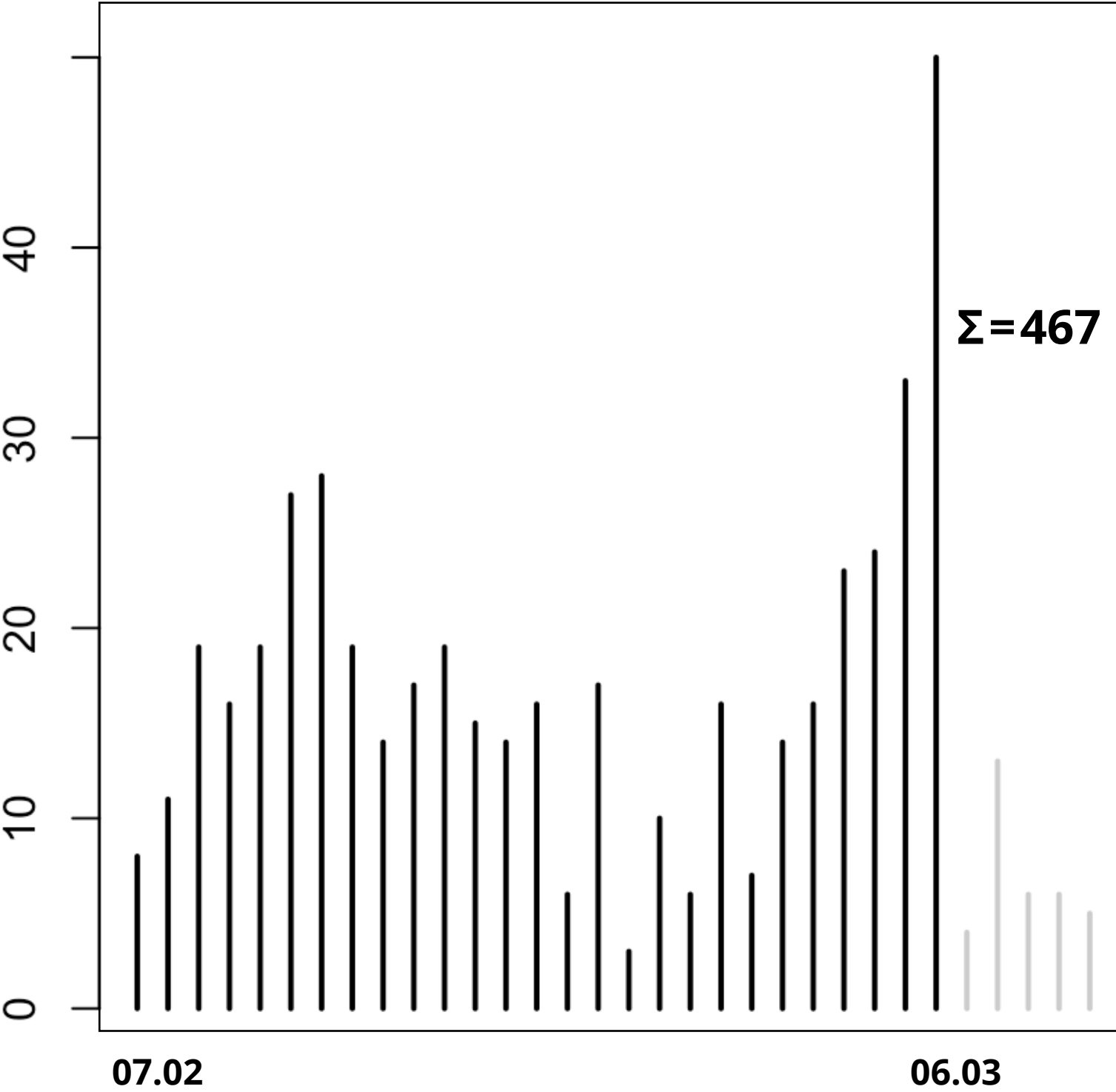


Турниры

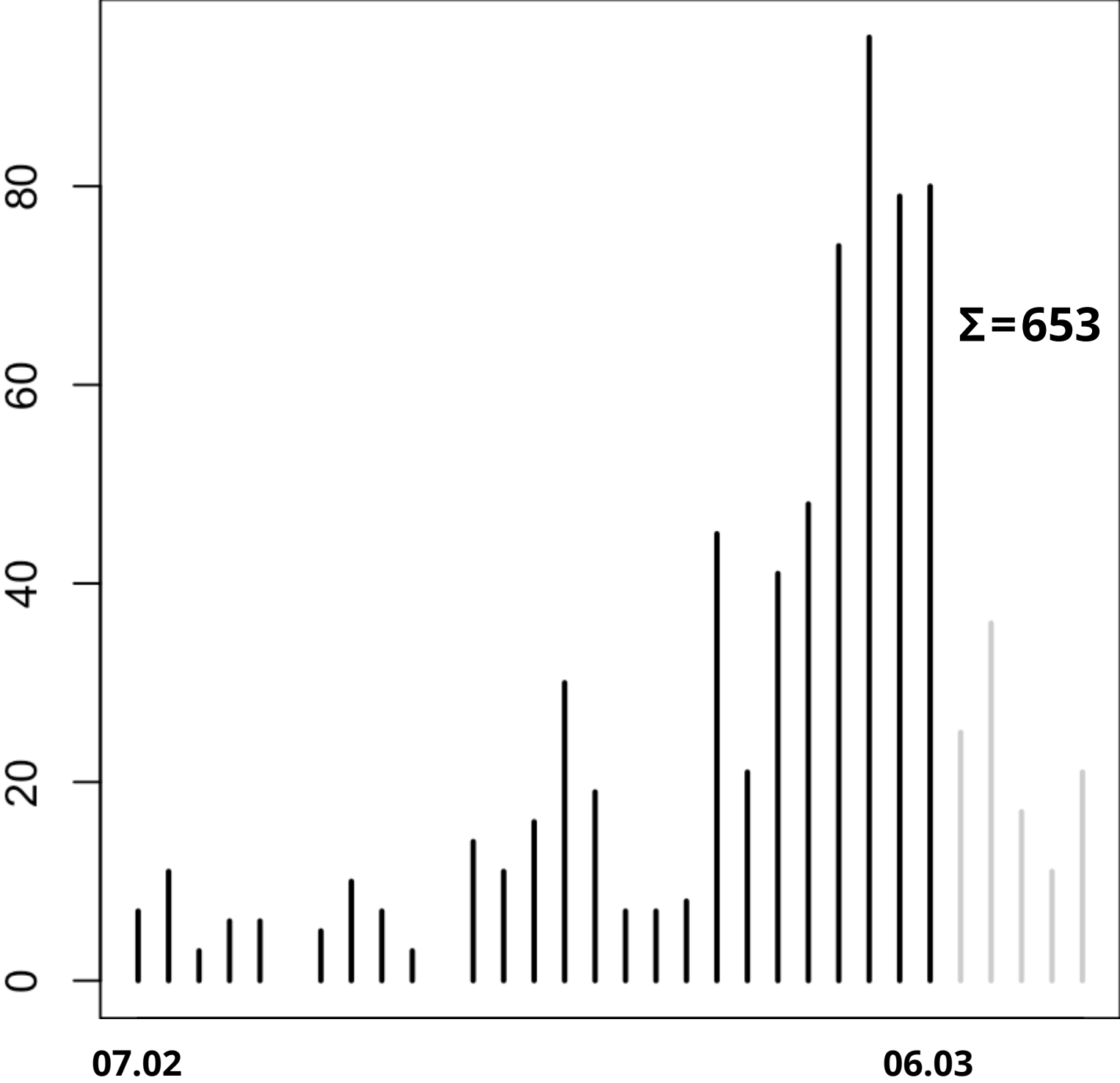


Сабмиты задач по первой фазе

Атака



Защита



Квалификация на первый турнир

Атака

#	Teams (45)		ROC-AUC Diff [?]	Last	Entries
1	Лаборатория	   	0.418524	6d	24
2	CONS		0.418017	9d	7
3	TransactionGPT		0.38832	7d	20
4	Gennady Shtech		0.376008	5d	12
5	Alexander Dudin		0.365216	6d	19
6	SFTB	 	0.346422	4d	13
7	overfit		0.345505	7d	14
8	ssh1		0.33968	9d	2
9	Andrey Kuznetsov		0.337958	6d	18
10	Anton Morozov		0.313267	4d	11

Защита

#	Teams (32)		Mean Harm ROC-AUC [?]	Last	Entries
1	Andrey Kuznetsov		0.747895	6d	44
2	Drastamat Kazanchian		0.740048	5d	19
3	Лаборатория	   	0.737348	6d	27
4	Andrei Kopylov		0.73649	26h	26
5	TransactionGPT		0.735322	6d	44
6	Timofey Zheleznyakov		0.734479	13h	33
7	CONS		0.733449	7d	32
8	Evgeny Kazenov		0.732744	6d	10
9	Kag le		0.731145	7d	9
10	Gennady Shtech		0.729354	5d	41

итоги турнира будут дальше...

Публичные решения по первой фазе

Атака

- 1. Решение ShT3ch, [github](#)
- 2. Решение Aaasenin, [github](#)
- 3. Решение Kdimon15, [github](#)

Защита

- 1. Решение KonderLip, [github](#)
- 2. Решение Blinov-89, [github](#)
- 3. Решение Mitrofanov-m, [github](#)
- 4. Решение Chebax, [kaggle](#)



Data Fusion Contest 2023 Chat
135 members

ShT3ch/Data-Fusion-2023-Attack



1 Contributor 0 Issues 0 Stars 0 Forks

aaasenin/Data-Fusion-2023-Attack-Companion



Простые изменения бейзлайна, улучшающие качество на лб с 0,044603 до 0,146785

1 Contributor 0 Issues 0 Stars 0 Forks

KonderLip/data-fusion2023-defence



Open solution for defence task from Data Fusion Contest 2023

1 Contributor 0 Issues 0 Stars 0 Forks

blinov-89/data_fusion_protect_2023



Data Fusion Contest 2023 — Task 2 Защита

1 Contributor 0 Issues 0 Stars 0 Forks

mitrofanov-m/vtb-data-fusion-2023-defence



Open solution for VTB ML-competition

0 Contributors 0 Issues 0 Stars 0 Forks

kdimon15/data_fusion_baseline_20...



Baseline for Data Fusion Contest 2023 "Attack"

1 Contributor 0 Issues 0 Stars 0 Forks

DATA

FUSION

CONTEST

Часть 2 — companion

2023

Победители Companion в первой фазе

Атака

Gennady Shtekh 

Привет. Вот решение **атаки** с градиентами  и матаном . Матана там одна строчка, дальше манипуляции с данными. Заменяв константу в одной строчке, можно получить **0.34**. Прямо в текущем виде, если просто запустить, даст 0.14. На моей A100 работает 5 минут, можно успеть засабмитить.

https://github.com/ShT3ch/Data-Fusion-2023-Attack/blob/main/attacck_publikc.ipynb

GitHub

Data-Fusion-2023-Attack/attacck_publikc.ipynb at main · ShT3ch/Data-Fusion-2023-Attack
Contribute to ShT3ch/Data-Fusion-2023-Attack development by creating an account on GitHub.

ShT3ch/**Data-Fusion-2023-Attack**



 1

Contributor

 0

Issues

 0

Stars

 0

Forks



Защита

Андрей Кузнецов

Первый весенний день, отличный повод опубликовать продвинутое открытое решение для задачи "Защита"



Генерируем новые признаки и обучаем **CatBoost** для улучшения качества и увеличения надежности банковской модели. Скор на public leaderboard: **0,720445**. Без лишней воды, полезные комментарии находятся в ноутбуке. Также есть файлы для отправки в тестирующую систему.

Ссылка на репозиторий: <https://github.com/KonderLip/data-fusion2023-defence>

GitHub

GitHub - KonderLip/data-fusion2023-defence: Open solution for defence task from Data Fusion Contest 2023

Open solution for defence task from Data Fusion Contest 2023 - GitHub - KonderLip/data-fusion2023-defence: Open solution for defence task from Data Fusion Contest 2023

KonderLip/**data-fusion2023-defence**



Open solution for defence task from Data Fusion Contest 2023

 1

Contributor

 0

Issues

 0

Stars

 0

Forks



Выступления ++

DATA

FUSION

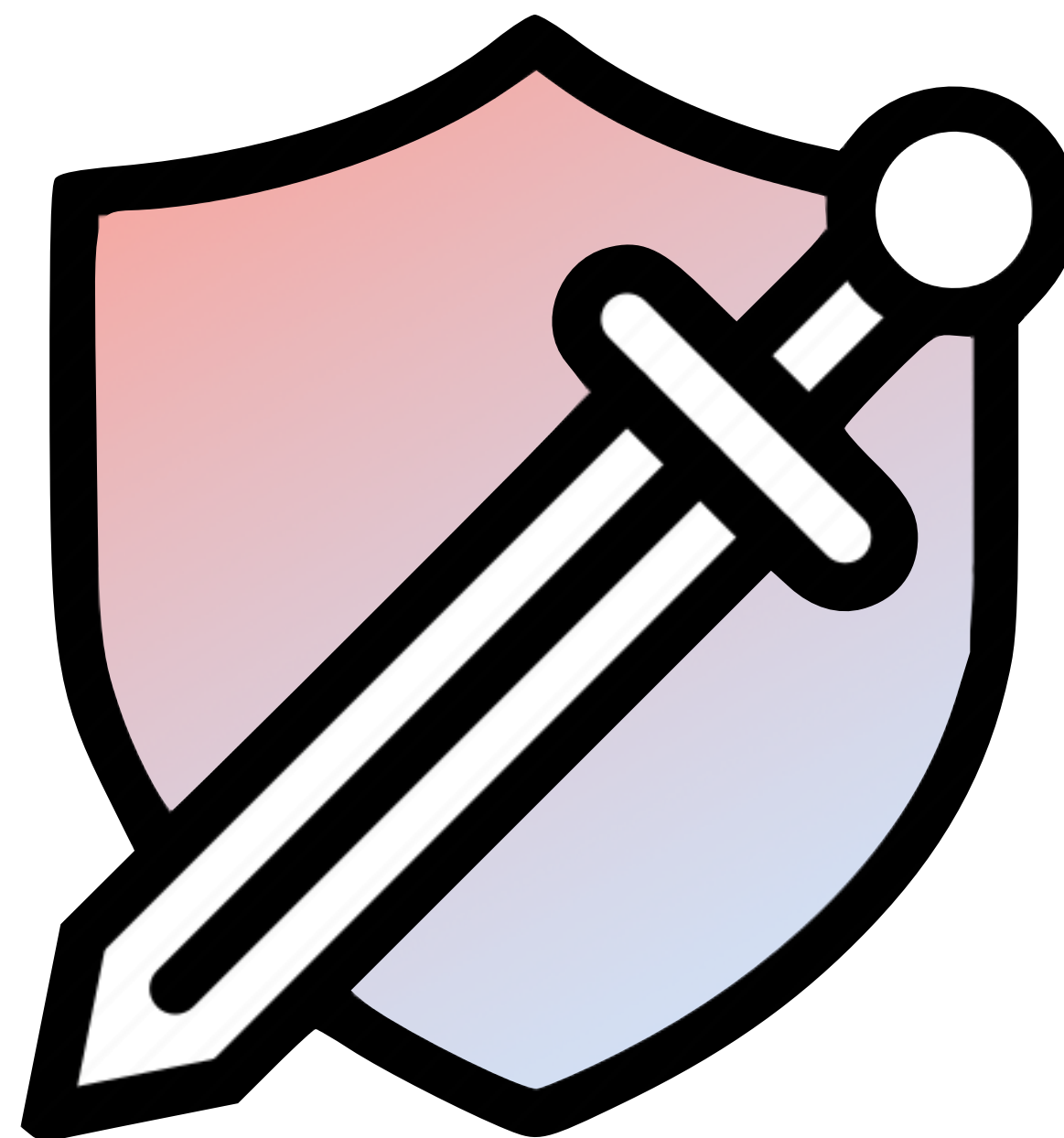
CONTEST

Часть 3 — турнир

2023

Первый (промежуточный) турнир

10 команд атаки
19 решений



10 команд защиты
20 решений

380 пар

Итоги первого турнира

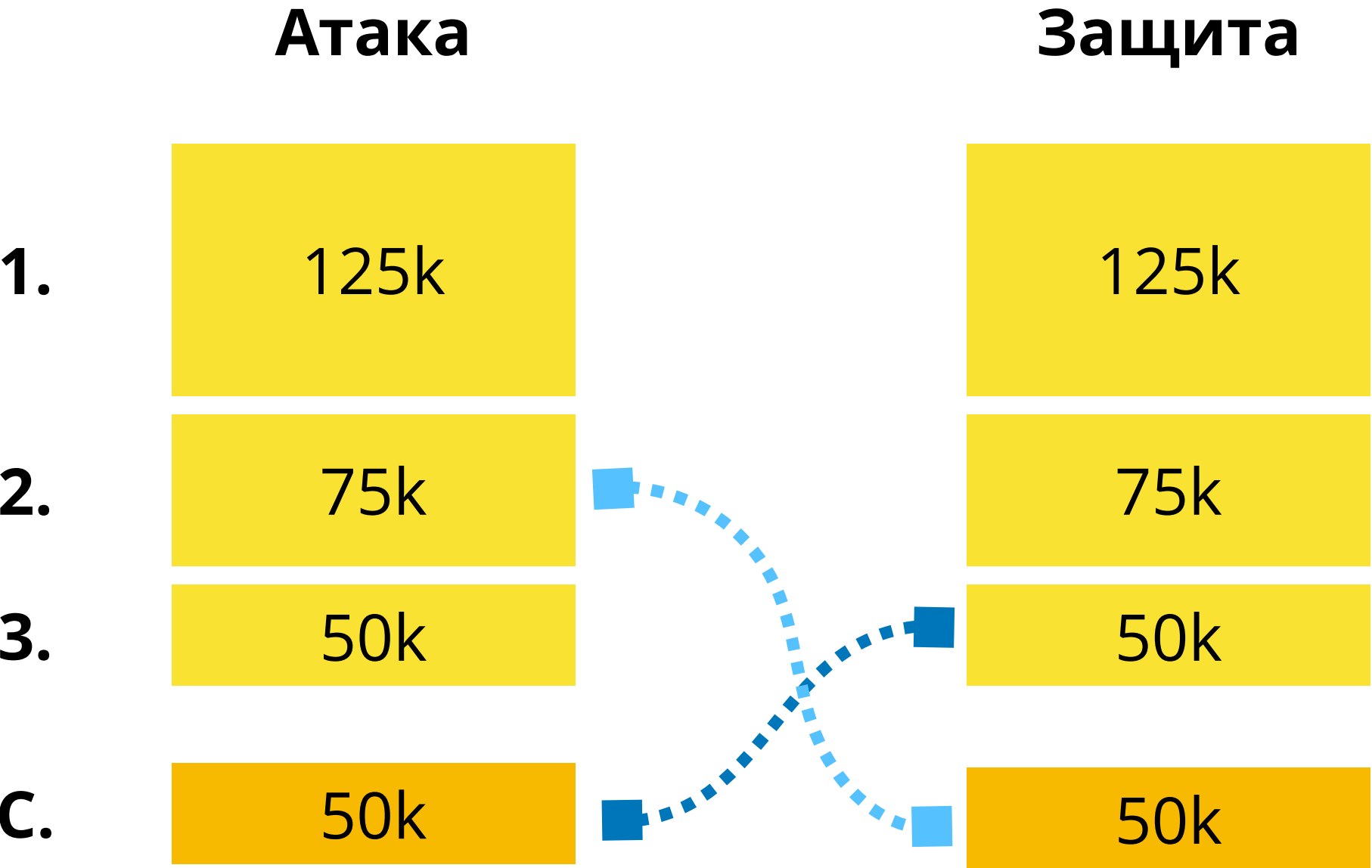
Атака

#	Teams (10)		ROC-AUC Diff [?]	Last	Entries
1	TransactionGPT		0.078472	7d	20
2	Andrey Kuznetsov		0.050208	6d	18
3	Alexander Dudin		0.025627	6d	19
4	Лаборатория	   	0.025113	6d	24
5	CONS		0.024203	9d	7
6	overfit		0.022697	7d	14
7	SFTB	 	0.020082	4d	13
8	Gennady Shtech		0.014378	5d	12
9	Anton Morozov		0.013327	4d	11
10	ssh1		0.00909	9d	2

Защита

#	Teams (10)		Mean Harm ROC-AUC [?]	Last	Entries
1	Timofey Zheleznyakov		0.688702	14h	33
2	Лаборатория	   	0.687847	6d	27
3	Gennady Shtech		0.68768	5d	41
4	CONS		0.686595	7d	32
5	TransactionGPT		0.666752	6d	44
6	Kag le		0.66414	7d	9
7	Andrei Kopylov		0.660798	27h	26
8	Andrey Kuznetsov		0.655214	6d	44
9	Drastamat Kazanchian		0.652665	5d	19
10	Evgeny Kazenov		0.636952	6d	10

Ирония, часть 1

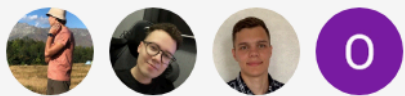


Ирония, часть 2 (самострел)

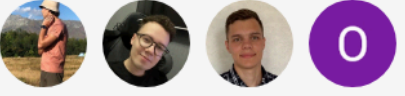


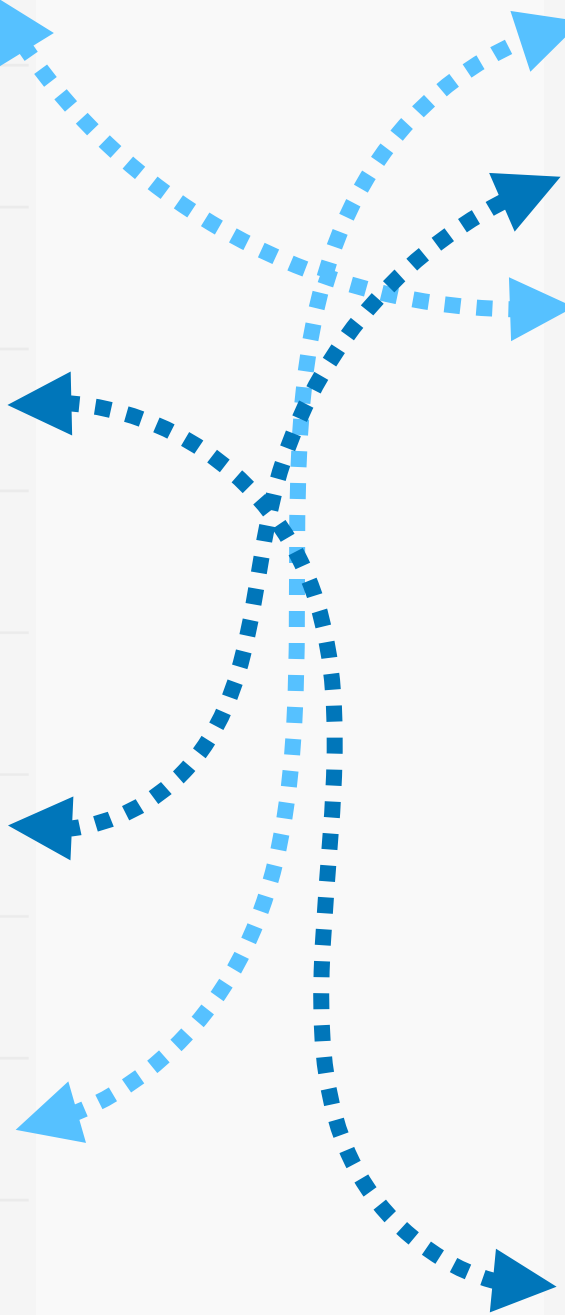
Квалификация на первый турнир (пары)

Атака

#	Teams (45)		ROC-AUC Diff [?]	Last	Entries
1	Лаборатория		0.418524	6d	24
2	CONS		0.418017	9d	7
3	TransactionGPT		0.38832	7d	20
4	Gennady Shtech		0.376008	5d	12
5	Alexander Dudin		0.365216	6d	19
6	SFTB	 	0.346422	4d	13
7	overfit		0.345505	7d	14
8	ssh1		0.33968	9d	2
9	Andrey Kuznetsov		0.337958	6d	18
10	Anton Morozov		0.313267	4d	11

Защита

#	Teams (32)		Mean Harm ROC-AUC [?]	Last	Entries
1	Andrey Kuznetsov		0.747895	6d	44
2	Drastamat Kazanchian		0.740048	5d	19
3	Лаборатория		0.737348	6d	27
4	Andrei Kopylov		0.73649	26h	26
5	TransactionGPT		0.735322	6d	44
6	Timofey Zheleznyakov		0.734479	13h	33
7	CONS		0.733449	7d	32
8	Evgeny Kazenov		0.732744	6d	10
9	Kag le		0.731145	7d	9
10	Gennady Shtech		0.729354	5d	41



Итоги первого турнира, поправка на самострел

Атака

#	Teams (10)		ROC-AUC Diff [?]	Last	Entries
1	TransactionGPT		0.078472	7d	20
2	Andrey Kuznetsov		0.050208	6d	18
3	Alexander Dudin		0.025627	6d	19
4	Лаборатория		0.025113	6d	24
5	CONS		0.024203	9d	7
6	overfit		0.022697	7d	14
7	SFTB		0.020082	4d	13
8	Gennady Shtech		0.014378	5d	12
9	Anton Morozov		0.013327	4d	11
10	ssh1		0.00909	9d	2

0.04424384

0.05347465

0.02385798

0.01381853

Защита

#	Teams (10)		Mean Harm ROC-AUC [?]	Last	Entries
1	Timofey Zheleznyakov		0.688702	14h	33
2	Лаборатория		0.687847	6d	27
3	Gennady Shtech		0.68768	5d	41
4	CONS		0.686595	7d	32
5	TransactionGPT		0.666752	6d	44
6	Kag le		0.66414	7d	9
7	Andrei Kopylov		0.660798	27h	26
8	Andrey Kuznetsov		0.655214	6d	44
9	Drastamat Kazanchian		0.652665	5d	19
10	Evgeny Kazenov		0.636952	6d	10

0.6877001

0.6948747

0.6547713

0.6523177

Гайды на будущее

1. Атакующим

- Проверьте что ваши файлы — разные
- Проверьте что файлы корректные
- Проверьте работоспособность кода

2. Защищающимся

- Не забудьте выбирать финальные решения
- Загружайте вовремя — сабмит “легализуется” после загрузки, а не с ее началом

3. Patchnote (1)

- В финальном турнире будет поправка на самострел

DATA

FUSION

CONTEST

Часть 4 — референдум

2023

Рефлексия

1. Три уровня скоров атаки

- White-box — 0.4+
- Black-box, самострел — 0.07+
- Black-box, IRL — 0.02 ... 0.05?

2. Устройство защиты

- Curated blackbox — атаки доподлинно неизвестны, но можно себя тестить
- Tradeoff — улучшать модель или наращивать защиту?
- Tradeoff — если улучшать, то насколько это устойчиво?

Adversarial **IRL**



Adversarial **IRL**



White-box
sandbox



Black-box
tournament



Ресар нашего предложения

1. Можем на вторую фазу положить

- Другие атаки для защиты
- Другую модель для атаки

2. ...При паре условий:

- Только публичные решения Фазы 1
- Решения должны быть не хуже организаторских
- За это проголосуют в telegram чате (да/нет)
- В чате выберут 1 модель и 1..3 атаки



Варианты изменений для второй фазы

1. Ничего не менять (white-box)
2. Подложить публичные решения ($\approx$ white-box)
3. Подложить решения победителей ($\approx$ black-box)
4. Подложить совершенно другие модели/атаки (black-box)

X

1. Для задачи Атака
2. Для задачи Защита

Обсуждение

- 1. Ничего не менять
- 2. Публичные решения
- 3. Решения победителей

X

- 1. Атака
- 2. Защита

Спасибо!

Telegram:

Data Fusion Contest 2023 Chat

